Amendments to the Claims:

This listing of claims replaces all prior versions and listings of claims in the application:

Listing of Claims:

1. (Currently amended) A method comprising:

receiving audio data including ~~having a~~ beat data;

extracting the ~~forming~~ beat data ~~based on~~ from said audio data;

determining a gesture window within which a gesture should occur, based on a specified time window relative to said beat data;

playing said audio data and obtaining video data during a time that said audio data is being played;

segmenting said video data to create a video clip ~~of~~ having a time ~~including~~ corresponding to the specified timing window; and

automatically determining ~~information related to~~ whether a predefined gesture ~~occurring~~ occurred in the video clip ~~only~~ within the specified timing window.

2. (Currently amended) The method of claim 1, wherein said determining includes determining a probability that each of ~~a~~

~~plurality of~~ one or more predefined gestures are performed within the timing window.

3. (Currently amended) The method of claim 2, wherein determining the probability that the video clip contains each of the predefined gestures includes evaluations of Hidden Markov Models.

4-6. (Canceled)

7. (Original) The method of claim 1, further comprising displaying a target gesture to be performed by the subject of the video data.

8. (Original) The method of claim 1, wherein each video clip contains video frames.

9. (Previously presented) The method of claim 8, further comprising identifying moving regions in each video frame in the video clip.

10. (Original) The method of claim 9, further comprising generating a feature vector for each video frame of the video clip.

11. (Previously presented) The method of claim 1, further comprising generating a score based on whether the video clip contains a target gesture.

12. (Original) The method of claim 11, further comprising displaying the score.

13. (Previously presented) The method of claim 11, wherein determining if the video clip contains the a target gesture includes generating a gesture probability vector having a plurality of elements, each element being associated with one of a plurality of predefined gestures and representing a probability that the video clip contains each of the associated predefined gestures.

14. (Currently amended) A system comprising:

an audio part~~,~~ to receive ~~receiving~~ audio data ~~having a~~ including beat data and ~~forming~~ extracting the beat data ~~based on~~ from said audio data;

a processor~~,~~ to determine ~~determining~~ a gesture window within which a gesture should occur, based on a specified time window relative to said beat data;

a temporal segmentor connected to receive video data during a time that said audio signal is being produced and to create a video clip from the video data ~~of~~, the video clip having a time ~~including said~~ corresponding to the specified time window; and

a recognition engine, in communication with the temporal segmentor, to determine if the video clip contains a predefined gesture, only within the specified timing window.

15. (Original) The system of claim 14, wherein the recognition engine includes a plurality of Hidden Markov Models.

16. (Previously presented) The system of claim 14, further comprising:

a video source, in communication with the temporal segmentor, to provide the video data to the temporal segmentor.

17. (Original) The system of claim 14, further comprising a move subsystem, in communication with the timing data source, to

provide a target gesture to be performed by the subject of the video data.

18. (Original) The system of claim 17, wherein the target gesture is a dance move that is to be performed by the subject of the video data.

19. (Original) The system of claim 17, further comprising a scoring subsystem, in communication with the recognition engine and the move subsystem, to determine if the video clip contains the target gesture.

20. (Original) The system of claim 19, further comprising a display subsystem, in communication with the scoring subsystem, to display a score that is a function of whether the video clip contains the target gesture.

21. (Original) The system of claim 20, wherein the display subsystem is in communication with the move subsystem and is configured to display a gesture request based on the target gesture.

22. (Original) The system of claim 14, wherein the recognition engine is configured to recognize predefined gestures and to produce a gesture probability vector having elements, each element being associated with one of the predefined gestures and representing the probability that the video clip contains the associated predefined gesture.

23-25. (Canceled)

26. (Currently amended) A computer program product, tangibly stored on a computer-readable medium, for recognizing gestures contained in video data, comprising instructions operable to cause a programmable processor to:

receive audio data including ~~having a~~ beat data;

~~form~~ extract the beat data ~~based on~~ from said audio data;

determine a gesture window within which a gesture should occur, based on a specified time window relative to said beat data;

obtain video data during a time that said audio signal is being produced;

segment said video data to create a video clip ~~of the~~ having a time ~~including said~~ corresponding to the specified timing window; and

automatically determine if the video clip contains a predefined gesture within the specified timing window.

27. (Canceled)

28. (Currently amended) An audio-visual processing system including:

a video source to provide video data;

an audio source to provide audio data ~~having a~~ including beat data;

a speaker to play at least a portion of the audio data; and

a computer program product, tangibly stored on a computer-readable medium, for recognizing gestures contained in video data, comprising instructions operable to cause a programmable processor, in communication with the video source and the audio source, to:

extract the beat data from the audio data;

determine a gesture window within which a gesture should occur, based on a specified time window relative to said beat data;

obtain video data during a time that said audio signal is being produced;

segment said video data to create a video clip based on said beat data; and

automatically determine if the video clip contains a predefined gesture ~~within only~~ within a specified timing window related to said beat data.

29. (Previously presented) The processing system of claim 28, wherein the computer program product further includes instructions operable to cause the programmable processor to:

perform a Hidden Markov Model process to determine if the video clip contains the predefined gesture.

30. (Previously presented) The processing system of claim 28, further comprising a display to display information based on whether the video clip contains the predefined gesture.